Version 1.0.0
July 5, 2010

Bayesian Reference Priors Analysis Code: `refpriors`

Luc Demortier[a], Supriya Jain[b] and Harrison B. Prosper[c] [1]

[a] *Laboratory of Experimental High Energy Physics, Rockefeller University*
[b] *Department of Physics, SUNY, Buffalo*
[c] *Department of Physics, Florida State University*

Abstract

The package `refpriors` allows one to construct Bayesian reference posterior distributions for cross section measurements and provides utilities to summarize these distributions via credibility intervals and Bayes factors. We provide a brief description of the C++ classes contained in this package and describe their use.

1 Introduction

This document describes the classes provided in the package `refpriors` and explains how to use them. The code implements the methods for computing Bayesian reference posteriors for cross section measurements discussed in Ref. [1], and includes examples illustrating their application. The document is structured as follows. Section 2 provides a brief description of the different classes and examples of their usage. Section 3 explains how to set up the package. Details on the algorithms used for calculating reference posteriors as well as examples with results are discussed in Sections 4 and 5. Our conclusions are contained in Section 6.

2 Description of the package

The `refpriors` package consists of the following main sub-directories:

- *src*: This contains a set of classes coded in C++ .
- *include*: This contains the headers for the above classes.
- *test*: This contains representative examples that illustrate the use of the above classes. Any additional code specific to a user's analysis can be added here (we call this the "user-code").

[1] luc.demortier@mail.rockefeller.edu, sjain@fnal.gov, harry@hep.fsu.edu

- *lib*: This contains the shared library built from the code in the *src* directory.
- *bin*: This contains the executables for the code in the *test* directory.
- *doc*: This area contains the files for this document.

2.1 Classes

We describe below briefly the different classes available in the *src* directory. The *test* directory contains examples of the use of some of these classes.

- *src/ReferencePrior*: This class implements José Bernardo's numerical algorithm for computing reference priors [2]. It can be used with *any* likelihood function, binned or unbinned, that depends on a single parameter. See Section 4 for a description.
- *src/JeffreysPrior*: This class implements Jeffreys' rule for calculating priors for any likelihood function that depends on a single parameter. As explained in Section 4, Jeffreys' rule coincides with the reference prior when some regularity conditions are met.
- *src/Binomial*: This class defines the binomial probability mass function (PMF), see Section 4.1 for details. It serves to illustrate the methods for defining a PMF that are needed to interface properly with *ReferencePrior*, *JeffreysPrior*, and other classes. For example, *ReferencePrior* requires a method to sample data from the PMF and a method to return a likelihood value given observed data and a value for the parameter of interest.
- *src/Exponential*: This class defines the exponential probability density function (PDF).
- *src/Poisson*: This class defines the Poisson PMF.
- *src/PoissonGammaModel*: This class implements the multiple-bin Poisson-gamma model, so called because it combines a multiple-bin Poisson likelihood with gamma priors. The likelihood is a product of Poisson probability mass functions over one or more bins. In each bin i, the Poisson mean depends on the signal cross section σ, the bin-dependent background magnitude μ_i, and the bin-dependent product ϵ_i of the signal acceptance and the integrated luminosity. We sometimes refer to ϵ_i as the effective integrated luminosity. The conditional prior $\pi(\vec{\epsilon}, \vec{\mu}|\sigma)$ for the nuisance parameters μ_i and ϵ_i is a product of gamma densities. This model is used with the `BAT` package [3] to calculate the Method 1 reference prior described in Ref. [1] (see Section 5 for details).
- *src/MarginalizedPoissonGamma*: This class implements the Poisson-gamma model marginalized with respect to the nuisance parameters $\vec{\mu}$ and $\vec{\epsilon}$. It is used in the calculation of the reference prior for a cross section measurement using the Method 2 approach described in Ref. [1]. In this method the likelihood is first

integrated over the nuisance parameters, and Jeffreys' rule is then applied to the result (see Section 4 for details).

- *src/SingleBinPoissonGammaMethod2Prior*: This class computes an analytical expression for Jeffreys' prior for the single-bin Poisson-gamma model marginalized with respect to the nuisance parameters μ and ϵ. The reference prior coincides with Jeffreys' prior for this model.

- *src/SingleBinMarginalizedPoissonGammaB*: Here we consider a slightly modified single-bin Poisson-gamma model, in which the likelihood is Poisson with mean $s + \mu$ instead of $\epsilon\sigma + \mu$; the parameter s is the signal magnitude. This class marginalizes this modified Poisson-gamma model with respect to the background magnitude μ (see Appendix B). The modified Poisson-gamma model could be useful in a "model-independent" counting experiment where one does not wish to commit to a particular signal model, and hence to a signal-model-dependent value of ϵ.

- *src/SingleBinPoissonGammaBMethod2Prior*: This class computes an analytical expression for Jeffreys' prior for the single-bin modified Poisson-gamma model marginalized with respect to the parameter μ. The reference prior coincides with Jeffreys' prior for this model.

- *src/SingleBinPoissonGammaMethod1UpperLimit*: Computes upper limits on the signal cross section for the single-bin Poisson-gamma model, using the analytical formula for the Method-1 posterior given at the end of section III.A.1 in Ref. [1]. This class can also be loaded directly in ROOT. Thus for example, if one has an expected background of 0.23 ± 0.05, an effective integrated luminosity of 2.0 ± 0.4, and sees 4 events, the 95% credibility level (C.L.) upper limit can be obtained by invoking ROOT and typing:

  ```
  .L SingleBinPoissonGammaMethod1UpperLimit.cc
  SingleBinPoissonGammaMethod1UpperLimit(2.0,0.4,0.23,0.05,4,0.95);
  ```

  This should result in an upper limit of 4.52724, in the inverse of the units used to specify the effective integrated luminosity.

- *src/Posterior*: This class computes the Bayesian posterior density, given the prior as one of the inputs.

- *src/CentralIntervals*: Given the posterior density at given θ values, this class calculates the central interval at y C.L., where y is specified by the user.

- *src/Coverage*: Given a probability density function, a table of n observed counts, and C.L. intervals, this class calculates the coverage probability for a specified true value of the parameter of interest. Note this class is currently not used in any of the examples available in the *test* directory, but is provided as an additional utility.

- *src/BayesFactor*: Given the likelihoods for the background-only and signal + background hypotheses, and a predicted value and uncertainty for the signal, this class computes the Bayes factor in favor of the signal hypothesis and against the background-only hypothesis. A log-normal distribution is used for the signal prior. A comprehensive review of the history of Bayes factors, their interpretation, and applications to various problems is provided in Ref. [4].

- *src/SingleBinKullbackLeibler*: Given two single-bin probability mass functions $p(k\,|\,\theta)$ and $q(k\mid\phi)$ (that is, discrete distributions), each with a single, nonnegative parameter of interest, θ and ϕ respectively, this class computes the Kullback-Leibler (K-L) divergence between them. The K-L divergence, an invariant, nonnegative measure of the separation between two distributions, is zero if and only if the latter are identical. For discrete distributions, the K-L divergence is given by

$$\kappa(\phi\,|\,\theta) \;=\; \sum_{k=0}^{\infty} p(k\,|\,\theta)\ln\frac{p(k\,|\,\theta)}{q(k\,|\,\phi)}.$$

  For fully specified models, the K-L divergence is simply the expected log-likelihood ratio in favor of the model that generated the data [5]. More generally, it is an expected log-Bayes factor. If p and q are Poisson distributions with means θ and ϕ respectively, the K-L divergence is given by $\kappa(\phi\,|\,\theta) = -\theta+\phi+\theta\ln(\theta/\phi)$, which, writing $\phi=\mu$ and $\theta=s+\mu$, becomes $\kappa(\mu\,|\,s+\mu) = -s+(s+\mu)\ln(1+s/\mu)$. We can get some insight into the meaning of the K-L divergence by considering the limit in which the signal to background ratio is small, $s/\mu << 1$. In this limit

$$\kappa(\mu\,|\,s+\mu) = -s+(s+\mu)\ln\Big(1+\frac{s}{\mu}\Big) \approx -s+(s+\mu)\Big[\frac{s}{\mu}-\frac{1}{2}\frac{s^2}{\mu^2}+\cdots\Big] \approx \frac{1}{2}\frac{s^2}{\mu},$$

  that is, $\sqrt{2\kappa(\mu\,|\,s+\mu)}$ reduces to $s/\sqrt{\mu}$, a commonly used measure of signal "significance" [2].

- *src/SingleBinDiscrepancy*: Given two single-bin probability mass functions p and q, each with a single parameter of interest, this class computes the expected discrepancy

$$\delta(\phi,n) \;=\; \int_0^\infty d(\phi,\theta)\,p(\theta\,|\,n)\,d\theta \;=\; \frac{1}{p(n)}\int_0^\infty d(\phi,\theta)\,p(n\,|\,\theta)\,\pi_R(\theta)\,d\theta,$$

  between them, where $d(\phi,\theta) = \min[\kappa(\phi\,|\,\theta),\kappa(\theta\,|\,\phi)]$ is called the *intrinsic discrepancy* [2], $p(n) = \int_0^\infty p(n\,|\,\theta)\,\pi_R(\theta)\,d\theta$, and $\pi_R(\theta)$ is the reference prior for the parameter of interest θ. This measure of discrepancy is invariant with respect to one-to-one transformations of the model parameters.

- *src/Utils*: This class contains some utilities used in the other classes.

- *src/Plot*: This class contains methods for making general purpose plots in ROOT.

[2] In this limit, the K-L divergence is symmetrical in its arguments, that is, $\sqrt{2\kappa(\mu\,|\,s+\mu)} = \sqrt{2\kappa(s+\mu\,|\,\mu)}$. In this case, the quantity $s/\sqrt{\mu}$ can be interpreted as a *distance* between the two models in the space of models.

2.2 Test programs

We describe briefly the test programs ("user-code") in the *test* directory.

- *test/testBinomial.cc*: This is user code for testing the reference prior algorithm on a binomial likelihood. It produces the plot Binomial.eps (Figure 1, left panel), see section 4.1 for details.
- *test/testExponential.cc*: This is user code for testing the reference prior algorithm on an exponential likelihood. It produces the plot Exponential.eps (Figure 1, middle panel), see section 4.1 for details.
- *test/testPoisson.cc*: This is user code for testing the reference prior algorithm on a Poisson likelihood. It produces the plot Poisson.eps (Figure 1, right panel), see section 4.1 for details.
- *test/testPGammaMethod2.cc*: This is user code for testing the reference prior algorithm on a marginalized Poisson-gamma likelihood, following Method 2 in Ref. [1]. It produces the plots PGamma.eps (left panel of Figure 2) and PGammaJeffreysPrior.eps (right panel of the same figure), see section 5.1 for details.
- *test/testPGammaMethod1Method2.cc*: This is user code that shows how to calculate the reference priors using Methods 1 and 2 from Ref. [1]. It also demonstrates the use of the priors in obtaining posterior densities, as well as central intervals and Bayes factors. It produces the plots testBAT.eps, testPGamma12Prior.eps, and testPGamma12Posterior.eps, all shown in Figure 3, see section 5.2 for details.
- *test/testPGammaBMethod2.cc*: This is user code for testing the Method-2 reference prior algorithm on a marginalized Poisson-gamma likelihood without the effective integrated luminosity parameter (see Appendix B). It produces the plot PGammaB.eps (Figure 4).
- *test/testDiscrepancy.cc*: User code for testing the class SingleBinDiscrepancy.
- *test/KullbackLeibler.cc*: User code for testing the class SingleBinKullbackLeibler.

3 How to set up the package

The `refpriors` package uses the ROOT [6] and BAT [3] utilities. Hence users must add these two packages to their working directory prior to using the `refpriors` package, and specify their paths appropriately. The ROOT and BAT packages can be downloaded from `http://root.cern.ch/drupal/` and `http://www.mppmu.mpg.de/bat` respectively. For testing our implementation of the `refpriors` code we have used the following versions of ROOT and BAT:

- ROOT: 5.22/00 (gcc 3.4)
- BAT: 0.3.2.

Once the paths for the ROOT and BAT utilities have been appropriately set, the `refpriors` code may be compiled and run from the main `refpriors` directory as follows:

- To compile, type: *make*.
  This will create the shared library of all code in the "lib" directory, and the executables of all the user-code in the "bin" directory.
- To run a specific user-code, for example "test/testBinomial.cpp",
  type: *./bin/testBinomial*.

Results from the examples are created in the main `refpriors` directory.

4 Reference priors for one-parameter models

Consider a one-parameter probability density function $p(x \,|\, \theta)$, where x is the dataset and θ an unknown parameter. The reference prior for θ is then defined as [2]:

$$\pi_R(\theta) \;=\; \lim_{k\to\infty} \frac{\pi_k(\theta)}{\pi_k(\theta_0)},$$

with

$$\pi_k(\theta) \;=\; \exp\left\{ \int p(x_{(k)} \,|\, \theta)\, \ln\left[\frac{p(x_{(k)} \,|\, \theta)\, h(\theta)}{\int p(x_{(k)} \,|\, \theta)\, h(\theta)\, d\theta} \right] dx_{(k)} \right\}, \tag{4.1}$$

where θ_0 is an arbitrary fixed value of θ, $h(\theta)$ is any continuous, strictly positive function, such as $h(\theta) = 1$, and $p(x_{(k)} \,|\, \theta) = \prod_{i=1}^{k} p(x_i \,|\, \theta)$ is the likelihood for k independent observations. In general, the analytical derivation of reference priors can be extremely challenging. However, as shown by Bernardo [5], (4.1) is readily amenable to numerical integration using the following pseudo-code. Let θ be a parameter value at which it is desired to evaluate the reference prior $\pi_R(\theta)$:

- Starting values:
  Choose a moderate value for k (to simulate the asymptotic posterior).
  Choose an arbitrary positive function $h(\theta)$, say $h(\theta) = 1$.
  Choose the number M of samples to be simulated.
- For the given value of θ:

```
1  Set P = 0
2  Repeat the following steps, for m = 1, ..., M:
3      Generate a random sample x_1, ..., x_k of size k from the PDF p(x | θ)
4      Evaluate Q_m = ∫ [p(x_(k) | θ') h(θ') dθ'] / [p(x_(k) | θ) h(θ)]
5      Set P = P + ln Q_m
6  Set π_R(θ) = exp(−P/M).
```

The one-dimensional integration at line 4 is done numerically.

A further useful result mentioned in Ref. [1] is that, when certain regularity conditions are met — essentially those that guarantee asymptotic normality of the posterior

— the reference prior for models with one continuous parameter reduces to the well-known Jeffreys prior [7]:

$$\pi_J(\theta) \;=\; \sqrt{\mathbb{E}\left[-\frac{d^2}{d\theta^2}\ln p(x\,|\,\theta)\right]}, \tag{4.2}$$

where the expectation is taken with respect to the sampling model $p(x\,|\,\theta)$. For simple models it may be possible to compute Jeffreys' prior exactly, but with increasing model complexity an analytical calculation may not be feasible. In such cases, (4.2) can be numerically approximated:

$$\pi_J(\theta) \;\approx\; \sqrt{-\frac{1}{N}\sum_{i=1}^{N}\frac{d^2}{d\theta^2}\ln p(x_i\,|\,\theta)}. \tag{4.3}$$

Here, $\{x_i\}$ is a sample of size N generated from $p(x|\theta)$. If the 2^{nd} order derivatives are too cumbersome to calculate analytically, they can be computed numerically, as explained in Appendix A.

4.1 Examples

We consider three simple models for which the reference prior coincides with Jeffreys' prior and is known in closed form:

- *Binomial*: The available data sample consists of n Bernoulli trials (with n fixed in advance), of which x are successes. The model is Binomial:

$$p(x\,|\,n,\theta) \;=\; \binom{n}{x}\theta^x\,(1-\theta)^{n-x}, \qquad \text{for } x=0,\ldots,n \text{ and } 0<\theta<1. \tag{4.4}$$

  The Jeffreys' prior for this model is known exactly:

$$\pi(\theta) \;=\; \sqrt{\theta^{-1}\,(1-\theta)^{-1}}. \tag{4.5}$$

- *Exponential*: The data follow an exponential PDF:

$$p(x\,|\,\theta) \;=\; \theta\,e^{-\theta x}, \qquad \text{for } x>0 \text{ and } \theta>0. \tag{4.6}$$

  The Jeffreys' prior for this model is:

$$\pi(\theta) \;=\; \frac{1}{\theta}. \tag{4.7}$$

- *Poisson*: Here the data follow a Poisson PDF:

$$p(x\,|\,\theta) \;=\; \frac{e^{-\theta}\,\theta^x}{x!}, \qquad \text{for } x>0 \text{ and } \theta>0. \tag{4.8}$$

  The Jeffreys' prior for this model is:

$$\pi(\theta) \;=\; \frac{1}{\sqrt{\theta}}. \tag{4.9}$$

For each of the above models we compute the reference prior using the numerical reference algorithm described in Section 4, and compare it to the appropriate analytical expression for Jeffreys' prior given above, as well as to a numerical approximation based on equation (4.3). Figure 1 shows the results — we see excellent agreement between the different calculations for all three models.

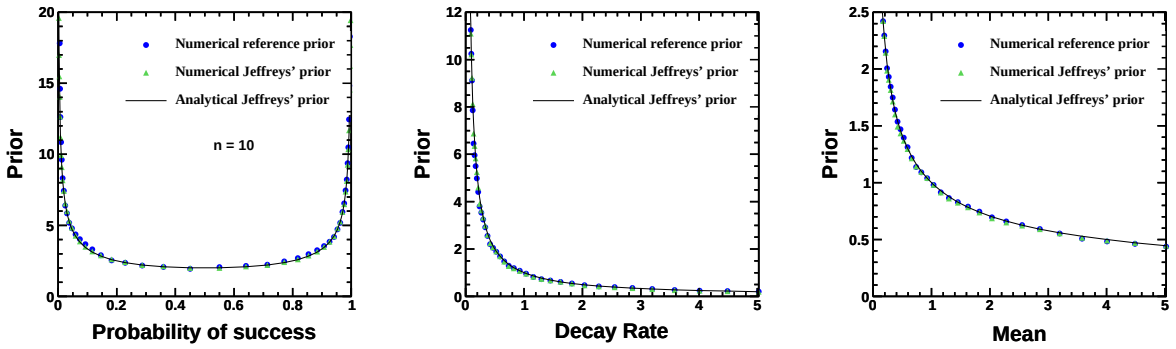

Figure 1: Application of Bernardo's Monte Carlo algorithm to the calculation of reference priors for binomial (left), exponential (middle), and Poisson (right) models. For the binomial model the population size is set to $n = 10$. In each plot, the analytical formula for Jeffreys' prior (solid line) is compared with Bernardo's numerical integration algorithm (circles) and with a numerical differentiation algorithm based on eq. (4.3) (triangles).

5 Reference priors for models with nuisance parameters

The construction of reference priors for one-parameter models can be generalized to models with nuisance parameters about which partial information is available. We first look at the "single-bin" or "single-channel" case, and then generalize to the multiple-bin case.

5.1 Single-bin models

A very common model for high energy physics measurements is the following [1]. A number of events N is observed by some apparatus, and it is assumed that N is Poisson distributed with mean count $\epsilon\,\sigma + \mu$, where σ is the rate of a physics signal process, typically the cross section, which we detect with an effective integrated luminosity ϵ — that is, the integrated luminosity scaled by the signal efficiency, and μ is a background contamination. Thus, σ is the parameter of interest, whereas ϵ and μ are nuisance parameters for which we usually have partial information. For physical reasons none of these three parameters can be negative. We write the likelihood for this model as

$$p(n|\sigma,\epsilon,\mu) \;=\; \frac{(\epsilon\sigma+\mu)^n}{n!}\,e^{-\epsilon\sigma-\mu} \quad \text{with} \;\; 0 \le \sigma < \infty \;\; \text{and} \;\; 0 < \epsilon, \mu < \infty. \tag{5.1}$$

Information about ϵ and μ usually comes from a variety of sources, such as auxiliary measurements, Monte Carlo simulations, theoretical calculations, and evidence-based beliefs (for example, some sources of background contributing to μ may be deemed small enough to ignore, and some physics effects on ϵ, such as gluon radiation, may be believed to be well enough reproduced by the simulation to be reliable "within a factor of 2"). It is therefore natural to represent that information by an evidence-based prior. Here we will assume that ϵ and μ are independent of σ and that their prior factorizes as a product of two gamma densities:

$$\pi(\epsilon, \mu \,|\, \sigma) \;=\; \pi(\epsilon, \mu) \;=\; \frac{a(a\epsilon)^{x-1/2}\, e^{-a\epsilon}}{\Gamma(x+1/2)} \, \frac{b(b\mu)^{y-1/2}\, e^{-b\mu}}{\Gamma(y+1/2)}, \tag{5.2}$$

where a, b, x, and y are known constants.

For this model, there are two plausible ways to calculate the reference prior for σ while incorporating the effect of nuisance parameters:

Method 1: Compute the conditional reference prior $\pi_R(\sigma \,|\, \epsilon, \mu)$ for the interest parameter σ given a fixed value of (ϵ, μ); the full prior is then $\pi(\sigma, \epsilon, \mu) = \pi_R(\sigma \,|\, \epsilon, \mu)\, \pi(\epsilon, \mu)$;

Method 2: Marginalize the probability model $p(n|\sigma, \epsilon, \mu)$ with respect to (ϵ, μ) in order to obtain $p(n|\sigma) = \int p(n|\sigma, \epsilon, \mu)\, \pi(\epsilon, \mu|\sigma)\, d\epsilon\, d\mu$, and compute the reference prior $\pi_R(\sigma)$ for the marginalized model; the full prior is then $\pi(\sigma, \epsilon, \mu) = \pi(\epsilon, \mu \,|\, \sigma)\, \pi_R(\sigma)$.

The advantage of Method 1 is that it produces a reference prior that does not need to be recomputed every time the form of the evidence-based prior is changed. Method 2, on the other hand, has the advantage that it reduces the problem to that of finding the reference prior for a one-parameter model, allowing the user to apply the methods of section 4. This approach is illustrated in Figure 2, where Method 2 is applied to the Poisson-gamma model with $x = y = a = b = 25$. The left panel compares the result of Bernardo's reference algorithm, a numerical evaluation of Jeffreys' prior, and an analytical calculation described in section III.A.2 of Ref. [1]. Good agreement is observed. The right panel compares two numerical calculations of Jeffreys' prior. The one labeled (1) is based on equation (4.3) (with n_i replacing x_i and σ replacing θ), whereas the one labeled (2) uses:

$$\pi_J(\sigma) \;=\; \sqrt{\mathbb{E}\left\{\left[\frac{d}{d\sigma} \ln p(n \,|\, \sigma)\right]^2\right\}}, \;\approx\; \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{d}{d\sigma} \ln p(n_i \,|\, \sigma)\right]^2}, \tag{5.3}$$

where the n_i are generated from $p(n \,|\, \sigma)$. The equality of these two expressions for Jeffreys' prior is derived in Ref. [8], among others.

5.2 Multiple-bin models

The multiple-bin model can be obtained by considering M replications of the single-bin one. The likelihood is:

$$p(\vec{n} \,|\, \sigma, \vec{\epsilon}, \vec{\mu}) \;=\; \prod_{i=1}^{M} \frac{(\epsilon_i \sigma + \mu_i)^{n_i}}{n_i!}\, e^{-\epsilon_i \sigma - \mu_i}, \tag{5.4}$$

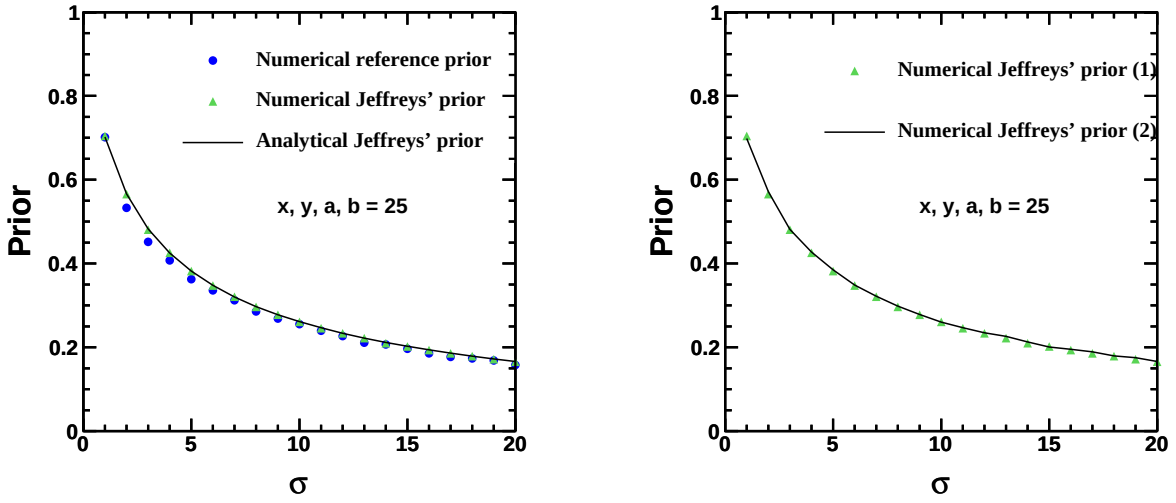

Figure 2: Left: Method-2 calculation of the reference prior for the Poisson-gamma model using Bernardo's algorithm (circles), a numerical approximation of Jeffreys' prior (triangles), and an analytical formula (solid line). Right: comparison of two numerical calculations of Jeffreys' prior for the same model (see text).

and the evidence-based prior is:

$$\pi(\vec{\epsilon},\vec{\mu}|\sigma) \;=\; \pi(\vec{\epsilon},\vec{\mu}) \;=\; \prod_{i=1}^{M} \frac{a_i(a_i\epsilon_i)^{x_i-1/2}\,e^{-a_i\epsilon_i}}{\Gamma(x_i+1/2)}\;\frac{b_i(b_i\mu_i)^{y_i-1/2}\,e^{-b_i\mu_i}}{\Gamma(y_i+1/2)}. \tag{5.5}$$

The generalization of Methods 1 and 2 to the multiple-bin model is straightforward. Numerical techniques for computing the corresponding reference priors are described next.

For Method 1 the algorithm starts by generating $(\sigma,\vec{\epsilon},\vec{\mu})$ triplets from the "flat-prior posterior", i.e. the posterior obtained by setting $\pi(\sigma\,|\,\vec{\epsilon},\vec{\mu}) = 1$ (line 3 in the pseudo-code below); the correct reference prior $\pi(\sigma\,|\,\vec{\epsilon},\vec{\mu})$ is then computed at lines 4–6 and is used at line 7 to weight the generated σ values so as to produce the reference posterior:

1 Set $\vec{n}_o$ to the array of observed event numbers.
2 For $i = 1,\ldots,I$:
3 Generate $(\sigma_i,\vec{\epsilon}_i,\vec{\mu}_i)$ from the probability model $p(\vec{n}_o\,|\,\sigma,\vec{\epsilon},\vec{\mu})\,\pi(\vec{\epsilon},\vec{\mu})$.
4 For $j = 1,\ldots,J$:
5 Compute the conditional Jeffreys' prior $\pi_J(\sigma_i\,|\,\vec{\epsilon}_i,\vec{\mu}_i)$ for σ_i, given the corresponding set of $(\vec{\epsilon}_i,\vec{\mu}_i)$, and using $p(\vec{n}\,|\,\sigma_i,\vec{\epsilon}_i,\vec{\mu}_i)$ for the probability model.
6 Histogram the σ_i values generated at line 3, weighting them by $\pi_J(\sigma_i\,|\,\vec{\epsilon}_i,\vec{\mu}_i)/p(\vec{n}_o\,|\,\sigma_i,\vec{\epsilon}_i,\vec{\mu}_i)$. This yields $\pi_{R1}(\sigma)$, the σ-marginal prior.
7 Histogram the σ_i values generated at line 3, weighting them by $\pi_J(\sigma_i\,|\,\vec{\epsilon}_i,\vec{\mu}_i)$. This yields $\pi_{R1}(\sigma\,|\,\vec{n}_o)$, the σ-marginal posterior.

Although not required for the calculation of the reference posterior, an approximation to the reference prior is provided at line 6. By construction this approximation is

only reliable for σ values in the bulk of the flat-prior posterior. The generation step at line 3 is done via a Markov chain Monte Carlo (MCMC) procedure [3]. The particular choice of sampling distribution for the generated set of $(\sigma, \vec{\epsilon}, \vec{\mu})$ is motivated by the desire to obtain weights with reasonably small variance at steps 6 and 7. However, the flat-prior posterior $p(\vec{n}_0 \,|\, \sigma, \vec{\epsilon}, \vec{\mu})\, \pi(\vec{\epsilon}, \vec{\mu})$ is not always proper with respect to $(\sigma, \vec{\epsilon}, \vec{\mu})$. When $M = 1$ (single-count model), it is improper if $x \leq 1/2$. Propriety can then be restored by multiplying the flat-prior posterior by ϵ and correspondingly adjusting the weights at steps 6 and 7. It may also be noted that if one histograms the σ_i values for each $i = 1, \ldots, I$, one obtains the distribution for $p(\vec{n}_o \,|\, \sigma)$. This can be compared to the exact calculation, which is the same as that derived in (5.8) below, and hence provides a test for the robustness of the MCMC procedure for generating the parameter values in line 3 above (see left panel of Fig. 3).

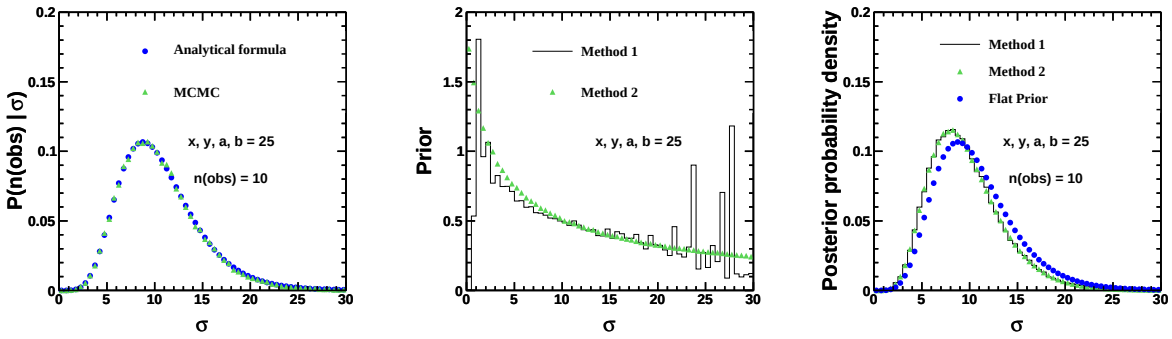

Figure 3: Calculation of reference posteriors for the single-bin Poisson-gamma model. Left: verification of the Markov chain Monte Carlo procedure in BAT used for sampling the set of parameters $(\sigma, \vec{\epsilon}, \vec{\mu})$. Middle: reference priors from Methods 1 and 2. Right: reference posteriors from Methods 1 and 2; also shown is the posterior for a flat prior.

The algorithm for Method 2 has a simpler structure, since all it does is apply Jeffreys' rule to a marginalized likelihood $p(\vec{n}_o \,|\, \sigma)$:

$$p(\vec{n} \,|\, \sigma) \;=\; \int_0^\infty d\vec{\epsilon} \int_0^\infty d\vec{\mu}\; p(\vec{n} \,|\, \sigma, \vec{\epsilon}, \vec{\mu})\, \pi(\vec{\epsilon}, \vec{\mu}). \tag{5.6}$$

If we make the following identifications for the multiple-count model ($i = 1, \ldots, M$):

$$\begin{aligned} A_i^1 &= x_i - 1/2, \\ A_i^2 &= y_i - 1/2, \\ p_i^1 &= \sigma / a_i, \\ p_i^2 &= 1/b_i, \end{aligned} \tag{5.7}$$

then (5.6) can be written as [9]:

$$p(\vec{n} \,|\, \sigma) \;=\; \prod_{i=1}^{M} p(n_i|\sigma) \;=\; \prod_{i=1}^{M} \left[\sum_{k=0}^{n} C_i^{k,1} C_i^{n-k,2} \right], \tag{5.8}$$

where

$$\begin{aligned} C_i^{0,j} &= (1+p_i^j)^{-(A_i^j+1)}, \\ C_i^{r,j} &= \left(\frac{p_i^j}{1+p_i^j}\right)\left(\frac{A_i^j+r}{r}\right) C_i^{r-1,j}, \quad (r=1,\ldots,n,\ j=1,2). \end{aligned} \tag{5.9}$$

The calculation for reference priors in Method 2 does not require random sampling of the parameters and is done at fixed σ values. For a given σ, the reference prior $\pi_{R2}(\sigma)$ is obtained from a numerical approximation of the Jeffreys' prior given the probability model of a marginalized Poisson-Gamma (5.8). The reference posterior is proportional to the product of $p(\vec{n}_o \,|\, \sigma)$ and $\pi_{R2}(\sigma)$, and the normalization with respect to σ must be determined numerically by requiring $\int p(\sigma \,|\, \vec{n}_o)\, d\sigma = 1$.

The above algorithms can, of course, also be applied to single-bin models. This is illustrated in Figure 3 for the single-bin Poisson-gamma model. The priors and posteriors obtained using Methods 1 and 2 are shown in the middle and right panels, respectively. For comparison, the posterior density for a flat prior distribution is also shown. The jaggedness of the Method 1 prior reflects the fluctuations due to the MCMC sampling of the parameters. It is more pronounced in regions of σ where the sampling distribution, shown in the left panel, has low density.

6 Conclusions

We have provided a brief description of the code available in the package `refpriors`, and details of representative examples illustrating the use of the code. We hope this document will allow the user to get started with the available software. Users may add their own examples in the *test* directory, and even define new probability models in the *src* directory, depending on their particular physics needs.

Appendix

A Numerical approximation of 2^{nd} order derivatives

If the 2nd derivatives are too cumbersome to be done analytically, they can be computed numerically, using, for example, the approximation

$$\frac{d^2y}{dx^2} = D(y,h)/h^2 - \frac{h^2}{12}\frac{d^4y}{dx^4} + \bigcirc(h^4), \tag{A.1}$$

where the difference operator is defined by $D(y,h) = y(x+h) - 2y(x) + y(x-h)$. Equation A.1 follows from the Taylor expansions of $y(x+h)$ and $y(x-h)$ about x. If we neglect the term of $\bigcirc(h^2)$ we get a formula accurate to that order.

We can get a much more accurate formula by computing Eq. (A.1) with a step size of h and a coarser one of size $2h$

$$\frac{d^2y}{dx^2} = D(y,2h)/(2h)^2 - 4\frac{h^2}{12}\frac{d^4y}{dx^4} + \bigcirc(h^4). \tag{A.2}$$

The $\bigcirc(h^2)$ term can be canceled by summing Eqs. (A.1) and (A.2) with weights 4/3 and -1/3, respectively, to arrive at

$$\frac{d^2y}{dx^2} = [4D(y,h)/3 - D(y,2h)/12]/h^2 + \bigcirc(h^4), \tag{A.3}$$

which is accurate to 4th order in h. Note, because the difference formula requires the computation of $y = \ln p(n|\sigma - 2h)$, and given that $\sigma \geq 0$, the smallest value of σ that can be computed is $\sigma = 2h$.

B Single bin $s + \mu$ Poisson-gamma model

A special case of the Poisson-gamma model is obtained by writing $s = \epsilon\sigma$ and dropping the prior on ϵ. This model describes a counting experiment in which the parameter of interest is not the cross section but rather the signal magnitude s. For this model, the marginal likelihood is

$$\begin{aligned} p(n \,|\, s) &= \int p(n \,|\, s, \mu)\; \pi(\mu \,|\, s)\, d\mu, \\ &= \int \frac{(s+\mu)^n}{n!}\, e^{-s-\mu}\, \frac{b(b\mu)^{y-1/2}}{\Gamma(y+1/2)}\, e^{-b\mu}\, d\mu, \\ &= \left[\frac{b}{b+1}\right]^{y+\frac{1}{2}} \sum_{k=0}^{n} v_{nk}\, \frac{s^k}{k!}\, e^{-s}, \end{aligned} \tag{B.1}$$

where

$$v_{nk} \equiv \binom{y - \frac{1}{2} + n - k}{n-k} \left[\frac{1}{b+1}\right]^{n-k}, \tag{B.2}$$

and for non-integral x,

$$\binom{x+m}{m} \equiv \frac{\Gamma(x+m+1)}{\Gamma(x+1)\,m!}. \tag{B.3}$$

The associated Method-2 reference prior is

$$\pi_{R2}(s) \;\propto\; \sqrt{e^{-s}\sum_{n=0}^{\infty}\frac{[T_n^0 - T_n^1/s]^2}{T_n^0}}, \tag{B.4}$$

where

$$T_n^m \;\equiv\; \sum_{k=0}^{n} k^m\, v_{nk}\,\frac{s^k}{k!} \quad \text{for} \quad m=0,1. \tag{B.5}$$

The subscript $R2$ refers to Method 2.

B.1 Algorithm for numerical calculation

The calculation of $p(n|s)$ and T_n^m is most easily done using the recursive functions:

$$\begin{aligned}
W_0(s,z) &= 1, \quad \text{for } n>0 \text{ and } e^{-s} \text{ for } n=0,\\
W_k(s,z) &= z\left(\frac{s}{k}\right) W_{k-1} \quad \text{for} \quad k=1,\cdots,n,\\
Y_0(z) &= 1,\\
Y_k(z) &= z\left(\frac{y-\frac{1}{2}+k}{k}\right)\left(\frac{1}{b+1}\right) Y_{k-1}, \quad \text{for} \quad k=1,\cdots,n.
\end{aligned} \tag{B.6}$$

We can then write

$$p(n\,|\,s) \;=\; \left[\frac{b}{b+1}\right]^{y+\frac{1}{2}} \sum_{k=0}^{n} W_k(s,z)\, Y_{n-k}(z), \tag{B.7}$$

with $z=e^{-s/n}$ for $n>0$ and $z=1$ for $n=0$, and

$$T_n^m \;=\; \sum_{k=0}^{n} k^m\, W_k(s,1)\, Y_{n-k}(1). \tag{B.8}$$

Figure 4 compares the analytical result (B.4) with Bernardo's algorithm and Jeffreys' prior (equation 4.3). The three calculations agree.

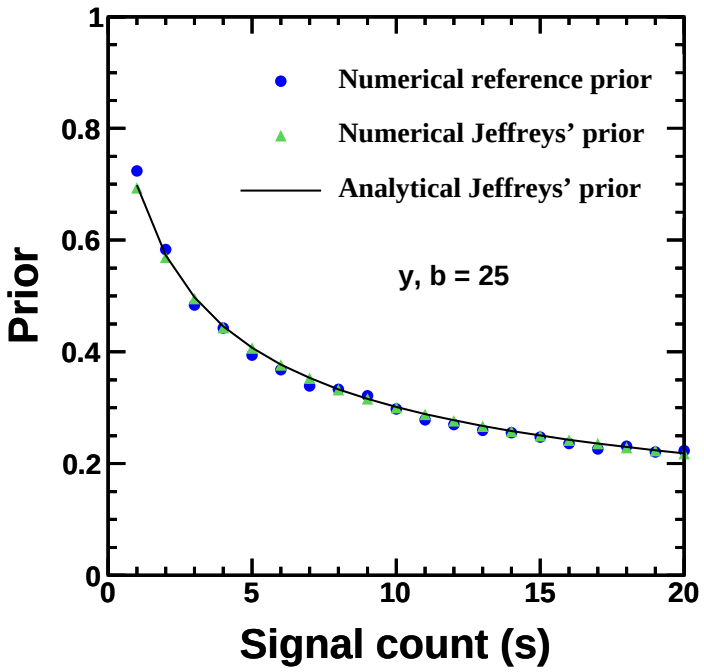

Figure 4: Method-2 calculations of the reference prior for the Poisson-gamma model without the effective integrated luminosity parameter: Bernardo's algorithm (circles), a numerical approximation of Jeffreys' prior (triangles), and an analytical formula (solid line).

References

[1] L. Demortier, S. Jain, H. B. Prosper, "Reference Priors for High Energy Physics," http://arxiv.org/abs/1002.1111. 1, 2, 3, 5, 6, 8, 9

[2] J. O. Berger, J. M. Bernardo, and D. Sun, "The formal definition of reference priors," Ann. Statist. **37**, 905 (2009); http://www.uv.es/~bernardo/2009Annals.pdf. 2, 4, 6

[3] A. Caldwell, D. Kollar, K. Kröninger, "BAT - The Bayesian Analysis Toolkit," Comp. Phys. Comm. **180**, 2197 (2009); http://arxiv.org/abs/0808.2552; http://www.mppmu.mpg.de/bat. 2, 5, 11

[4] R. E. Kass and A. E. Raftery, "Bayes factors," J. Amer. Statist. Assoc. **90**, 773 (1995); http://www.stat.washington.edu/raftery/Research/PDF/kass1995.pdf. 4

[5] J. M. Bernardo, "Reference analysis," in *Handbook of Statistics 25*, edited by D. K. Dey and C. R. Rao (Elsevier, Amsterdam, 2005), pp. 17-90; http://www.uv.es/~bernardo/RefAna.pdf. 4, 6

[6] http://root.cern.ch. 5

[7] R. E. Kass and L. Wasserman, "The selection of prior distributions by formal rules," J. Amer. Statist. Assoc. **91**, 1343 (1996); http://www.stat.cmu.edu/~kass/papers/rules.pdf. 7

[8] See for example section 17.14 in A. Stuart, K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics, Volume 2A: Classical Inference and the Linear Model* (Edward Arnold, London, 1999). 9

[9] P. C. Bhat, H. B. Prosper, and S. S. Snyder, "Bayesian Analysis of Multi-Source Data," Phys. Lett. **B** 407, 73 (1997); http://dx.doi.org/10.1016/S0370-2693(97)00723-5. 11